Predicting the Future of Car Manufacturing Industry using Data Mining Techniques

Dr. M.Hanumanthappa[1] , Sarakutty.T.K [2]
[1] Department of Computer Science & Applications, Bangalore University, Bangalore, India
hanu6572@hotmail.com
[2] Department of Computer Science & Applications, Dayananda Sagar College, Bangalore, India
sarajobythomas@yahoo.co.in

***Abstract*—Data mining helps to find predictive information from large databases. Companies use predictive modeling tools for strategic decision-making. It helps companies to identify and account for the key assumptions that drive business value—enabling good decision making that leads to predictable results. By analyzing the company's historical information we can anticipate these changes. This paper aims at providing a proposed data mining solution that can be used for automotive market, especially in the car manufacturing domain. That is to predict the future sales on the base of historical data. Especially we aim at finding the number of cars to be manufactured by a car manufacturing company by using the previous years data. For this linear regression analysis technique is used.**

***Index Terms*—Data Mining, Predictive Modeling, Linear regression Analysis, WEKA**

I. Introduction

One of the largest and fastest growing industries in the global market is automobile. Being the leader in product and process technologies in the manufacturing sector, it has been recognized as one of the drivers of economic growth. For realizing the sector's full potential, well directed efforts have been made to provide a new look to the automobile policy [1]. The automobile industry is the largest industry in the world with revenues of about 1.8 trillion US dollars. Some call it the 'industry of industries'. The industry is rightfully engaged in building new plants, increasing production from existing plants, bringing out new models, exporting more and captivating users in innovative ways. According to Economist Intelligence Unit by 2020 almost 40 percent of the car sales will be in Asia, and the production of car components will shift to emerging markets of India and China. With more than 2 million new automobiles rolling out each year, on Indian roads, the industry is set to grow further. Automobile industry made its silent entry in India in the nineteenth century. Since the launch of the first car in 1897, Indian automobile industry has come a long way. The Indian car industry has undergone tremendous change in recent times in terms of innovative designs, concepts and technology. There are some of the car manufacturing companies in India like Toyota, Hyundai, Maruti Suzuki, Ford and Skoda which hope to make it big in next few years. These and other major players such as M&M and Hindustan Motors are devising new techniques to accelerate growth. Many of them have already launched the new luxury to economic class car models in India and many others are in the process to tap the market with their new offerings. The latest figures issued by SIAM showed that the Indian car market was a witness to a growth of 32.69% last year (Jan-Dec 2010). India is soon becoming a hub for car manufacturers not only to sell their cars but to setup manufacturing units. India comes a close second to China when it comes to the fastest growth in the Automobile sector in the world. Some of the challenges which the future car industry's are going to face are the less loyal customers, the need to improve productivity, the demand for producing low cost mass market vehicles and the maintenance of small top-end vehicle market. This paper aims to predict the future of car manufacturing company by using data mining techniques. Especially we aim at finding the number of cars to be manufactured by a car manufacturing company by using the previous years data. A formula based on the current data available, historical trends, and projections is used to estimate the total number of cars to be produced in a particular year.

II. Data mining

Data mining techniques are the result of a long process of research and product development. This evolution began when business data was first stored on computers, continued with improvements in data access, and more recently, generated technologies that allow users to navigate through their data in real time. Data mining takes this evolutionary process beyond retrospective data access and navigation to prospective and proactive information delivery. Data mining is ready for application in the business community because it is supported by three technologies Massive data collection, Powerful multiprocessor computers and Data mining algorithms. Data mining tools are used to predict future trends and behaviors. In order to make fact based decisions that have a positive effect on their business performance companies use Predictive Analytics and period Predictive Analytics uses analysis of current and historical data to predict future outcomes. The basic application of predictive analytics is to capture relationships in historical data to predict future outcomes, to guide decision making with minimum risks, searching the databases for hidden patterns, allowing businesses to make proactive, knowledge-driven decisions and to answer business questions that traditionally were too time consuming. Most companies already collect and refine massive quantities of data. Business intelligence organizations, financial analysts, healthcare management and medical diagnosis use data mining to extract information from the enormous data sets generated by modern experimental and observational methods [6] [7].

DOI: 01.IJIT.01.02.150

A. *Predictive Modelling*

Usually, data mining tasks can be categorized into either prediction or description [8]. Descriptive mining techniques are Clustering, Association Rule Mining (ARM) and Sequential pattern mining [9]. The predictive mining techniques are Classification, Regression and Deviation detection [10]. One of the most-used subfield of data mining is predictive modeling which is a combination of statistics, machine learning, database techniques, pattern recognition, and optimization techniques. Predictive modelling is a process used in predictive analytics to create a statistical model of future behaviour. Predictive analytics is the branch of data mining concerned with the prediction of future probabilities and trends. The central element of predictive analytics is the predictor, a variable that can be measured for an individual or other entity to predict future behaviour. Multiple predictors are combined into a predictive model, which, when subjected to analysis, can be used to forecast future probabilities with an acceptable level of reliability. In predictive modelling, data is collected for the relevant predictors, a statistical model is formulated, predictions are made and the model is validated (or revised) as additional data becomes available. The model may employ a simple linear equation or a complex neural network, mapped out by sophisticated software. Predictive analytics are applied to many research areas, including meteorology, security, genetics, economics, and marketing. Companies use predictive modelling tools for strategic decision-making. It helps companies identify and account for the key assumptions that drive business value—enabling good decision-making that leads to predictable results. In order to make the right decision we have to anticipate and plan for possible changes in the future. By analyzing the company's historical information we can anticipate these changes. To anticipate possible changes in the future, we must start addressing questions about the future possible outcomes, like Which ones are most likely?, Which ones matter most?, Which ones are best for the company? [2]

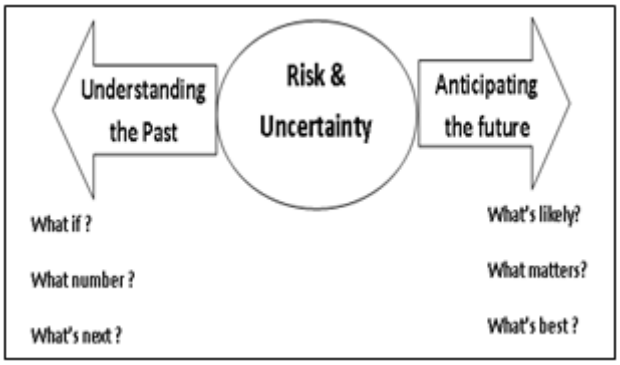

Figure 1 : Understanding the past can help model the future

III. Proposed methodology

Selecting a data mining algorithm is not an easy task, it depends upon: the data we have gathered, the problem we are trying to solve, and the computing tools that are available. Regression is the oldest and most well-known statistical technique that the data mining community utilizes. It is the most widely used method for numeric prediction. The relationship between one or more independent variables and dependent variable can be modelled using regression analysis [5]. The independent variable or predictor variable are the attributes of interest describing the tuple which are known. The dependent or response variable is what we want to predict. Regression analysis helps in predicting the value of the response variable, using predictor variables, whose values are already known. Basically, regression takes a numerical dataset and develops a mathematical formula that fits the data. In order to predict the future behaviour simply take the new data and plug it into the developed formula.

A. *Linear regression in car manufacturing domain*

In order to predict the future an understanding of the market, the trends, the moods, and the changing consumer tastes and preferences are required. Car manufacturing companies do not directly interact with the consumers. They rely upon the data that is provided by the market demands, vendors, previous years' data and the manufacturing capacity of their plants [4]. If a car manufacturing company has to decide on the number of cars to be manufactured in the next year, it has to do a thorough study on the market trends, the number of cars manufactured in the previous years, the number of cars sold in the previous years and also the manufacturing capacity of its plants. Linear regression is the simplest technique used to model continuous valued functions [5]. For prediction linear regression can be used to fit a predictive model to an observed data set of Y and X values. After developing such a model, if an additional value of X is then given without its accompanying value of Y, the fitted model can be used to make a prediction of the value of Y. Here we are trying to predict the number of cars to be manufactured by a car manufacturing company by considering the following numeric parameters. The response variable, Y = number of cars manufactured and Predictor variable, X = year. Now when an additional value for X that is a particular year is given it is possible to find the possible value for Y, the number of cars to be manufactured, using the following equations.

$$Y = b + wX \qquad (1)$$

Where 'b' is the regression coefficients specifying the 'Y' intercept 'w' is the regression coefficients specifying the slope of the line.

Now if we consider the regression coefficients as weights, then

$$Y = w_0 + w_1 X \qquad (2)$$

Let D be a training set consisting of values of predictor variable, X for some population and their associated values for response variable, Y. The training set contains |D| data points of the form (x1, y1), (x2, y2), ……., (x|D|, y|D|) [5].

The regression coefficients can be estimated with the method of least squares, were

$$w_1 = \frac{\sum_{i=1}^{|D|}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{|D|}(x_i - \bar{x})^2} \qquad (3)$$

DOI: 01.IJIT.01.02.150

and

$w_0 = \bar{y} - w_1\bar{x}$ (4)

Were $\bar{x}$ is the mean value of x_1, x_2 ,……., $x_{|D|,}$ and $\bar{y}$ is the mean value of y_1,y_2 ,……..,$y_{|D}$ [5].

B. Simulation and Numerical results

The sample data set was created by considering the previous years data. Using WEKA under the linear regression analysis classification technique this training set was executed. The predicted values for the classifiers were evaluated using 10-fold cross validation [11] and we have got the equation for the response variable. Using this linear regression on year equation it is possible to predict the number of cars to be manufactured in a particular year. The run information using WEKA is shown in figure 2 and the graph which shows the linear relationship between the two variables is plotted as in figure 3.

Figure 2: Run Information

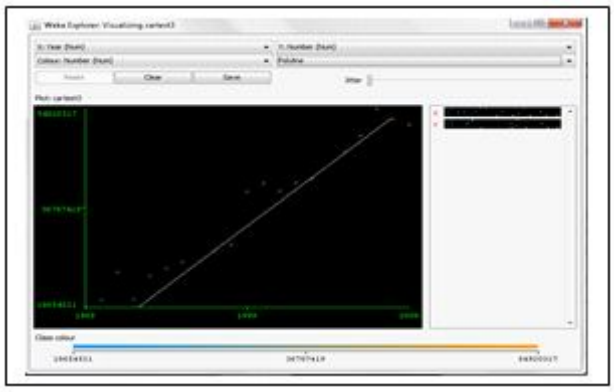

Figure 3: Graph

IV. CONCLUSION

This paper introduces the application of data mining technology in the car manufacturing unit and obtains an analysis result from small data, may expand the sample capacity in the practical application, to obtain more accurate conclusion. Some of the modifications that can be made to simple linear regression are rather than one predictor variable more predictor variables can be used. Transformations can be applied to the predictors, Predictors can be multiplied together and used as terms in the equation and modifications can be made to accommodate response predictions that just have yes/no or 0/1 values. Further work is under progress to develop an algorithm which can handle multiple predictor variables.

REFERENCES

[1] Abhijeet Kamble , Ajay Nalawade , Ashutosh Deo , Sagar Ranadive, "Demand/Sales Forecasting in Indian Firms", April 3, 2006.

[2] Driving Strategic Planning with Predictive Modeling: An Oracle White Paper.

[3] Alex Berson, Stephen Smith, and Kurt Thearling , "Building Data Mining Applications for CRM" .

[4] Mouhib Al-Noukari, Arab International University, Damascus, Syria & Wael Al-Hussan, The Arab Academy for Banking and Financial Sciences,Damascus, Syria, "Using Data Mining Techniques for Predicting Future Car market Demand : DCX Case Study ".

[5] Jiawei Han & Micheline Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers.

[6] David L. Iverson, "Data Mining Applications for Space Mission Operations System Health Monitoring", NASA Ames Research Center, Moffett Field, California, 94035, 2008.

[7] Ping Lu, Brent M. Phares, Terry J. Wipf and Justin D. Doornink, "A Bridge Structural Health Monitoring and Data Mining System", in Proceedings of the 2007 Mid-Continent Transportation Research Symposium, Ames, Iowa, August 2007.

[8] Tibebe Beshah Tesema, Ajith Abraham And Crina Grosan, "Rule Mining And Classification of Road Traffic Accidents Using Adaptive Regression Trees", In Proc. Of I. J. On Simulation, Vol. 6, No. 10 and 11, 2008.

[9] Chitriki Thotappa and Dr. Karnam Ravindranath " Data mining Aided Proficient approach for optimal inventory [1]
control in supply chain management" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 8, No. 2, May 2010

[10] S.Shankar, T.Purusothaman, "Utility Sentient Frequent Itemset Mining and Association Rule Mining: A Literature Survey and Comparative Study", International Journal of Soft Computing Applications, ISSN: 1453-2277 Issue 4 (2009), pp.81-95.

[11] Deepa Lakshmi .R and Radha .N, "Analyzing Data Mining Algorithms using Car Database", September 2009, DOI: DMKE092009001.

DOI: 01.IJIT.01.02.150